Tucker Laurence
Ryan Worley

The dataset we have chosen to explore for our project contains data from direct marketing campaigns by a Portuguese bank, conducted through phone calls. The purpose of the calls was to determine whether the client answered “yes” or “no” to a bank term deposit subscription. Often, multiple contacts with the same client were necessary.

The challenge lies in the inherent complexity of human behavior as reflected in the data. Clients' responses may be influenced by numerous factors, including demographics, previous contacts, and socio-economic indicators. Additionally, the requirement for multiple contacts introduces a temporal dimension, making the dataset more intricate. The high-dimensional feature space, coupled with the need for effective classification, poses an exciting challenge for machine learning algorithms. Furthermore, the dataset's size (41,188 instances in the full version) allows for robust model training, while the smaller datasets (4,119 instances) will enable the testing of computationally demanding algorithms like Support Vector Machines (SVM).

The dataset includes features on client information, such as age, job, marital status, and additional details like the method of contact, timing of the last contact, call duration, and more. Our target feature for classification is “has the client subscribed to a term deposit?” (binary: "yes" or "no"). We will analyze how the other features relate to this target feature and aim to build and train a machine learning model to classify clients’ responses as “yes” or “no” based on their information in the remaining features.

Tasks

1. Data Exploration and Preprocessing
    a. Load and explore datasets.
    b. Clean data
    c. handle missing values
    d. encode categorical variables.
2. Feature Selection and Engineering
    a. Identify significant features affecting subscription rates
    b. Engineer features if needed to improve modeling
3. Model Development
    a. Train a variety of models using different algorithms.
    b. Perform hyperparameter tuning using cross-validation.
4. Model Evaluation
    a. Evaluate models using train test and validation sets.
    b. Compare performance metrics across models.
5. Results Interpretation and Reporting
    a. Analyze results and interpret findings.
    b. Prepare a comprehensive report and presentation.
6. Final Review and Submission
    a. Finalize the report based on feedback.
    b. Submit the project.
7. Complete

We will both work together in at least some capacity in each step of the project to ensure that we both understand all the progress that is made, but certain steps will be “assigned” to each of us. For each step, it is the responsibility of the teammate assigned to that step to maintain the quality of the work being completed, and to ensure the work gets done on time. We will both contribute an equal amount to the project, and will hold each other accountable for doing our share of the work. To achieve this we will be in frequent communication over Microsoft Teams.

Timeline and Task Assignment:

Task	Dates	Time frame	Assignee
Finish and submit Proposal	10/30/2024 - 11/1/2024	3 days	Ryan
Data Exploration and Preprocessing	11/1/2024 - 11/4/2024	3 days	Tucker
Feature Selection and Engineering	11/4/2024 - 11/8/2024	4 days	Ryan
Model Development	11/8/2024 - 11/15/2024	1 week	Tucker
Model Evaluation	11/15/2024 - 11/22/2024	1 week	Ryan
Results Interpretation and Reporting	11/22/2024 - 11/29/2024	1 week	Tucker
Final Review and Submission	11/29/2024 - 12/6/2024	1 week	Ryan and Tucker
Complete	12/6/2024		